Read-Across or QSARs?
Which one to apply and when?

Presented by: Dr. Faizan SAHIGARA

Chemical Watch Expo 2017
26th April, 2017 | Berlin Germany

KREATiS, 23 rue du creuzat, 38080 L’Isle d’Abeau, France | Email: faizan.sahigara@kreatis.eu
Use of alternatives to animal testing under REACH

Reference: ECHA Practical guide – How to use alternatives to animal testing to fulfil your information requirements for REACH registration. Version 2.0 – July 2016
Read-Across is a technique for predicting endpoint information for one substance (target substance), by using data from the same endpoint from (an)other substance(s), (source substance(s)).

“QSAR is a mathematical model relating one or more quantitative parameters, which are derived from the chemical structure, to a quantitative measure of a property or activity”.

References:
ECHA Practical guide – How to use alternatives to animal testing to fulfil your information requirements for REACH registration. Version 2.0 – July 2016
ECHA Practical guide – How to use and report (Q)SARs version 3.1 – July 2016
An analogue approach: prediction for the target substance was based on available data for a very limited number of structurally similar substances.

Category approach: prediction for the target substance was based on available data on several structurally similar substances that are grouped together.

QSAR models are developed based on various statistical approaches including but not limited to:

- **a)** regression analysis
- **b)** classification methods,
- **c)** support vector machines,
- **d)** artificial neural networks.

Selection of approach depends on the complexity of the correlation between selected descriptors and the endpoint.

**Structural similarity**

- Patterns
- Breakdown products
- Data trends

**Metabolic pathways**

**Functional groups**

**Molecular descriptors**

**Model algorithms**

- Mode of action
- Applicability domain
- Mechanistic interpretation

**Read-Across vs. QSAR**
Read-Across vs. QSARs

Practical guide
How to use alternatives to animal testing to fulfil your information requirements for REACH registration
Version 2.0 – July 2016

Practical guide
How to use and report (Q)SARs
Version 3.1 – July 2016

Read-Across Assessment Framework (RAAF)

Guidance on information requirements and chemical safety assessment
Chapter F.6: QSARs and grouping of chemicals
May 2008

Guidance for the implementation of REACH
Read-across

- Category members are expected to be structurally similar or to follow a trend/pattern.

- For example, based on common functional groups, metabolites, chemical class, mode of action.

- Routes of exposure, duration of effects should also be taken into account.

- More analogues with reliable data indicates a strong category.

- Analogue approach by default is an extrapolation. Therefore, more justifications are expected.

- Category approach may also be associated with extrapolation provided the target is beyond its scope.

Read-across evaluation

Read-Across Assessment Framework (RAAF)

<table>
<thead>
<tr>
<th>SCENARIO</th>
<th>APPROACH</th>
<th>READ-ACROSS HYPOTHESIS BASED ON</th>
<th>QUANTITATIVE VARIATIONS</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Analogue</td>
<td>(Bio)transformation to common compound(s)</td>
<td>Property of the target substance predicted to be quantitatively equal to those of the source substance or prediction based on a worst-case approach.</td>
</tr>
<tr>
<td>2</td>
<td>Analogue</td>
<td>Different compounds have qualitatively similar properties</td>
<td>Properties of the target substance predicted to be quantitatively equal to those of the source substance or prediction based on a worst-case approach.</td>
</tr>
<tr>
<td>3</td>
<td>Category</td>
<td>(Bio)transformation to common compound(s)</td>
<td>Variations in the properties observed among source substances. Prediction based on a regular pattern or on a worst-case approach.</td>
</tr>
<tr>
<td>4</td>
<td>Category</td>
<td>Different compounds have qualitatively similar properties</td>
<td>Variations in the properties observed among source substances. Prediction based on a regular pattern or on a worst-case approach.</td>
</tr>
<tr>
<td>5</td>
<td>Category</td>
<td>(Bio)transformation to common compound(s)</td>
<td>No relevant variations in properties observed among source substances and the same strength predicted for the target substance.</td>
</tr>
<tr>
<td>6</td>
<td>Category</td>
<td>Different compounds have qualitatively similar properties</td>
<td>No relevant variations in properties observed among source substances and the same strength predicted for the target substance.</td>
</tr>
</tbody>
</table>

Most frequently encountered types of read-across approaches formulated as scenarios.

Characterised by scientific considerations, crucial to assess read-across (defined as assessment elements).

Includes logical order of questions and possible outcomes (defined as assessment options) and examples.

Evaluation of level of confidence and overall acceptability of the read-across approach.

Reference: ECHA Read-across assessment framework (RAAF). March 2017
Validity and reporting of QSARs

- Scientific validity of the model has been established.
- Target substance falls within the model’s applicability domain.
- QSAR prediction is fit for regulatory purpose.
- Information is well documented and as per REACH requirements.

5 OECD Principles for QSAR validation

<table>
<thead>
<tr>
<th>Principle</th>
</tr>
</thead>
<tbody>
<tr>
<td>A defined endpoint</td>
</tr>
<tr>
<td>An unambiguous algorithm</td>
</tr>
<tr>
<td>A defined domain of applicability</td>
</tr>
<tr>
<td>Appropriate measures of goodness-of-fit, robustness and predictivity</td>
</tr>
<tr>
<td>A mechanistic interpretation, if possible</td>
</tr>
</tbody>
</table>

QSAR documentation under REACH

- QSAR Model Reporting Format (QMRF)
  Justifies validity of QSAR model
- QSAR Prediction Reporting Format (QPRF)
  Justifies reliability in QSAR prediction

Reliable and well documented predictions from a valid model can be used as a stand-alone result.

Which approach to use and when?

Key deciding factors

a) Endpoint/Activity to predict (Physicochemical, Ecotoxicological, Human health)

b) Nature of endpoint (quantitative, qualitative)

c) Complexity of the endpoint

d) Availability of valid models with a defined applicability domain

e) Availability of reliable data for structurally similar substances

f) Structural complexity of the target substance (multiple functional groups, metabolism)

Which approach will be more acceptable to ECHA?

Which option will be more accurate?

Are QSARs less popular than read-across?

Which approach can completely replace experimentation?
The relevance of the alternative data: e.g. when using (Q)SAR, read across, categories or in vitro approaches, you should verify whether they are applicable for the substance (e.g. applicability domain of the (Q)SAR models, consistency of the category, relevance of the in vitro effects).

In practice, read-across is not encouraged for basic physico-chemical properties (e.g. water solubility, log Kow) since these properties provide key information for the assessment of a chemical in particular for the assessment of the environmental properties, and experimental data or valid QSAR predictions should normally be available (or should be reasonably obtainable).

It should be recognised that the robustness of a category approach would be expected to be considerably greater than that of an analogue approach, since the basis for evaluating any individual chemical in the category is greater, and there is usually more measured data available in such a wider approach.

Read-across from a negative result is regarded as equally valid and convincing as a positive result provided the test design, concentrations tested etc. have been chosen adequately.

Reference: ECHA Practical guide – How to use alternatives to animal testing to fulfil your information requirements for REACH registration. Version 2.0 – July 2016
ECHA Practical guide – How to use and report (Q)SARs version 3.1 – July 2016
ECHA guidance on information requirements and chemical safety assessment. Chapter R.6: QSARs and grouping of chemicals. May 2008
QSARs as supporting info to read-across results: Information from valid (Q)SARs may be used where possible to inform decisions on the need, extent and type of additional testing.

In general, you should use (Q)SAR results as part of a weight of evidence approach or an integrated testing strategy.

When using (Q)SARs, you should run all the available (Q)SAR models for the endpoint. The available models should be independent (different in terms of prediction formalism and underlying data).

ECHA’s experience of using adaptations to address standard informational requirements reveals that there are no simple (Q)SAR solutions for complex health endpoints such as repeated dose toxicity, developmental and reproductive toxicity in general.

Reference: ECHA Practical guide – How to use alternatives to animal testing to fulfil your information requirements for REACH registration. Version 2.0 – July 2016
ECHA Practical guide – How to use and report (Q)SARs version 3.1 – July 2016
ECHA guidance on information requirements and chemical safety assessment. Chapter R.6: QSARs and grouping of chemicals. May 2008
## Which approach to use and when? – example 1

<table>
<thead>
<tr>
<th>Endpoint name</th>
<th>Water solubility</th>
</tr>
</thead>
<tbody>
<tr>
<td>Endpoint type</td>
<td>Quantitative</td>
</tr>
<tr>
<td>Chemical type</td>
<td>Mono-constituent</td>
</tr>
<tr>
<td>Chemical category</td>
<td>Feasible with 10 category members (all &gt;80% structural similarity)</td>
</tr>
<tr>
<td>QSAR models</td>
<td>Available and some of them are scientifically valid for REACH purpose</td>
</tr>
<tr>
<td>Domain check</td>
<td>Yes within domain of the read-across category as well as some of the QSAR models</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Approach</th>
</tr>
</thead>
<tbody>
<tr>
<td>Read-across</td>
</tr>
<tr>
<td>QSAR</td>
</tr>
<tr>
<td>Both</td>
</tr>
<tr>
<td>None</td>
</tr>
</tbody>
</table>

QSAR should be sufficient provided the prediction is reliable, the model is scientifically validated using the 5 OECD principles and all documentation requested by ECHA was made available. If multiple valid models available, all of them should be considered. Most reliable can be used as key study. The rest as supporting.
### Which approach to use and when? – example 2

<table>
<thead>
<tr>
<th>Endpoint name</th>
<th>Skin Sensitisation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Endpoint type</td>
<td>Qualitative / Quantitative</td>
</tr>
<tr>
<td>Chemical type</td>
<td>Multi constituent</td>
</tr>
<tr>
<td>Chemical category</td>
<td>Analogue approach</td>
</tr>
<tr>
<td>QSAR models</td>
<td>Valid models available but cannot handle mixtures</td>
</tr>
<tr>
<td>Domain check</td>
<td>Extrapolation by default for analogue approach</td>
</tr>
</tbody>
</table>

**Read-across**

Read-across Analogue approach can be performed, however a strong justification should be used as basis. Metabolic pathways, hydrolysis/break-down products etc. should be analysed. RAAF requirements to be met. Perhaps, better to use as weight of evidence if no concrete justifications are provided.
### Which approach to use and when? – example 3

<table>
<thead>
<tr>
<th>Endpoint name</th>
<th>Mutagenicity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Endpoint type</td>
<td>Qualitative</td>
</tr>
<tr>
<td>Chemical type</td>
<td>UVCB (no clear info on composition)</td>
</tr>
<tr>
<td>Chemical category</td>
<td>5 mixtures with known composition available</td>
</tr>
<tr>
<td>QSAR models</td>
<td>Available but cannot handle UVCBs</td>
</tr>
<tr>
<td>Domain check</td>
<td>Inside category domain based on physicochemical properties</td>
</tr>
</tbody>
</table>

None of the approaches are reliable enough in this case to provide a quality prediction and conclude on the mutagenicity potential of the query substance.
### Which approach to use and when? – example 4

<table>
<thead>
<tr>
<th>Endpoint name</th>
<th>Acute aquatic toxicity to Fish (LC50)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Endpoint type</td>
<td>Quantitative</td>
</tr>
<tr>
<td>Chemical type</td>
<td>Multi-constituent</td>
</tr>
<tr>
<td>Chemical category</td>
<td>Feasible with 10 category members (all &gt;80% structural similarity)</td>
</tr>
<tr>
<td>QSAR models</td>
<td>Available and some of them are scientifically valid for REACH purpose</td>
</tr>
<tr>
<td>Domain check</td>
<td>Yes within domain of the read-across category as well as some of the QSAR models</td>
</tr>
</tbody>
</table>

QSAR should be sufficient provided the prediction is reliable, the model is scientifically validated using the 5 OECD principles and all documentation requested by ECHA was made available. Additionally, Quantitative Read-across prediction can be provided to support the QSAR prediction.
An overview of some Read-Across tools

1. OECD QSAR Toolbox

Latest version 4.0 was recently released

Category formation approach – primary basis for the tool

Data gap filling via Read-across, QSARs and trend analysis

Key features:

a) Endpoint relevant and empiric Profilers
b) Databases and inventories
c) Clearly defined workflow for data gap filling
d) Automated report generation for regulatory use
e) Metabolism simulators
f) Adverse Outcome Pathways (AOPs)

2. CEFIC-LRI AMBIT

Loaded with non-confidential REACH data supplied by ECHA

AMBIT database and functional modules allowing a variety of flexible searches and data mining.

Assessment tools for read-across and category formation

Examples of prediction tools incorporated into AMBIT:
- Cramer rules
- Protein binding
- Carcinogenicity and mutagenicity

Tools for data analysis:
- Regression, Classification, Clustering, etc.

Reference: https://ambitlri.ideaconsult.net/tool
More analogue searching tools listed in ECHA chapter R.6

An overview of some commonly used QSARs

1. Danish QSAR Database

Estimates from more than 200 (Q)SARs from free platforms (e.g. EPISuite) and commercial platforms (e.g. SciQSAR, LeadScope and others).

QSAR predictions for more than 600,000 chemical substances

Physicochemical properties, ecotoxicity, environmental fate, ADME and toxicity.

Key facts:

a) Consensus/battery prediction from multiple models
b) A clear indication of applicability domain status
c) Easy to generate report with all predictions
d) Free to access all incorporated free and commercial platforms.
e) Does not provide predictions for any user-defined structure (that is beyond those listed in its database).

Reference: http://qsar.food.dtu.dk/
2. VEGA

The QSAR models optimised in accordance with the REACH requirement.

Incorporates models from CAESAR or T.E.S.T., or those developed later by the contributors to VEGA.

QSAR models to predict tox, ecotox, environ, and phys-chem properties.

Key facts:

a) VEGAHub – allows access to all incorporated models and allows creating new user-defined models.

b) In silico models – VEGA, SarPY, CORAL, ToxRead, Janus, Prometheus, ToxWeight, ToxDelta.

c) Prediction report not necessarily sufficient under REACH. For example, no QPRF generated by the tool. Most structurally similar training set substances not necessarily reliable.

Reference: http://vegahub.eu/
An overview of some commonly used QSARs

3. EPISuite

EPI (Estimation Programs Interface) Suite™ Developed by US EPA and SRC.

Multiple models incorporated inside this tool: KOWWIN, WSKOWWIN, AEROWIN, ECOSAR etc.

Screening level tool

Key facts:

• Free to use and ideal for screening purposes

• Prediction reports available but does not provide QSAR documentation required under REACH

• No clear applicability domain evaluation for user-defined structures

Taking QSARs to a next level of precision

**iSafeRat® High Accuracy QSARs**

Developed by KREATiS, iSafeRat incorporates high precision prediction models for various physicochemical, ecotoxicological and human health endpoints.

Highly accurate predictions at a fraction of the price of the laboratory equivalent.

Study reports supported by the appropriate documents for regulatory acceptance (QMRFs and QPRFs, for REACH).

Interested to know more about KREATiS HA-QSARs and how we can help you?

We are exhibiting here along with our partner company CEHTRA. Visit us at our booth.
Structural alert systems – to support Read-Across and QSARs

**ToxTree**

Estimates toxic hazard by applying a decision tree approach.

17 plugins including: cramer rules, verhaar scheme, DNA and protein binding alerts etc.

Reference: http://toxtree.sourceforge.net/

**Nexus DEREK**

Expert, knowledge-based software to identify potentially toxic chemicals.

Provides detailed reasoning information about the likelihood of the toxicity in structure.

Reference: https://www.lhasalimited.org/products/derek-nexus.htm

Useful as screening tool. No alerts doesn’t necessarily indicate no toxicity.
More QSARs listed in ECHA practical guide

List is not complete. Doesn’t indicate other QSARs not accepted.

References:
ECHA Practical guide – How to use and report (Q)SARs version 3.1 – July 2016
CASE STUDIES
Substance identity:
Chemical name: cinnamaldehyde
CAS: 104-55-2
SMILES: O=CC=Cc1ccccc1

Profilers used:
Protein binding alerts for skin sensitisation by OASIS v1.3

Endpoint:
Skin sensitisation
Skin sensitisation ECETOC

Results:
LLNA EC3 = 3% (measured)
GPMT = Strong sensitizer (measured)

Despite the applicability domain issues in some cases, all QSAR models and read-across approach suggests the target substance is a skin sensitiser. Results from OECD QSAR Toolbox Skin Sensitisation AOP can be used a key result supported by reliable QSAR predictions.
Case Study 2: Skin irritation potential of thymol and p-thymol

Case study reflects how structurally similar substances can show a totally different skin irritation potential (one being corrosive, other non-irritant). Only iSafeRabbit HA-QSAR predicted both the substances accurately. Alert generated using OECD QSAR Toolbox can still be used to support iSafeRabbit prediction for thymol but this doesn’t hold true for p-thymol.
Case Study 3: Carcinogenicity potential of naphthalene

Substance identity:
Chemical name: naphthalene
CAS: 91-20-3
SMILES: c1ccccc2ccccc12
Experimental data: Positive results in both mouse and rat

<table>
<thead>
<tr>
<th>Model</th>
<th>Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>OECD QSAR Toolbox</td>
<td>Non-carcinogen (read-across)</td>
</tr>
<tr>
<td>VEGA</td>
<td>Inconclusive (reliability low)</td>
</tr>
<tr>
<td>Danish QSAR Database</td>
<td>Negative (liver specific cancer)</td>
</tr>
</tbody>
</table>

There is a low reliability and mismatching results between different in silico approaches. No consensus approach possible. None of the models or read-across here could be used on their own either.
Case Study 4: Water solubility of α-terpineol

Substance identity:
Chemical name: α-terpineol
CAS: 98-55-5
SMILES: CC1CCC(CC=1)C(C)(C)O
Exp WatSol: 2870 mg/L (ECHA, 2001, purge and trap 23°C, K2)

<table>
<thead>
<tr>
<th>Model</th>
<th>Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>iSafeRat HA-QSAR</td>
<td>2480 mg/L</td>
</tr>
<tr>
<td>EPISuite WSKOWWIN</td>
<td>371.1 mg/L</td>
</tr>
<tr>
<td>US EPA T.E.S.T</td>
<td>873.53 mg/L</td>
</tr>
</tbody>
</table>

As mentioned earlier, only QSAR prediction would be sufficient as it is not highly recommended to use Read-Across for physicochemical properties.
If you are interested to know more about KREATiS *in silico* services and how they can help you, do not hesitate to visit us at Booth 2.

Should you prefer to send in your queries by email: faizan.sahigara@kreatis.eu